

Text Summarization of Transcripts using NLP.

Abstract

With the advent of the COVID-19 pandemic, summarizing online recordings has become a more important task than ever before, with asynchronous timings of people involved in meetings sometimes meaning that they can't fully catch up. In this paper, we attempt to create a model that is able to take in long-form text, such as those found in lecture transcripts or presentation transcripts, and output an effective summary. We go over multiple approaches to this, using models such as BERT (Bidirectional Encoder Representations from Transformers), TextRank, and TF-IDF (Term Frequency - Inverse Document Frequency). From this, our results show that it is indeed possible to make an effective summarizer for high wordcount text, and show some interesting trends amongst the accuracy scores, with low precision and high recall that is ideal for summaries due to how the task of generating a summary is inherently different from predicting text.

Introduction

One of the most pressing problems with the advent of the COVID-19 pandemic has been the transition from offline mediums of communication to online interaction. To do so, many interfaces have been developed, most of which boil down to using online meeting spaces to coordinate meetings with people around the world. However, when setting up meetings with people in vastly different time zones, this comes with the problem that not everyone will be able to make it to the meetings. This is most evident in the case of online classes, where the large size of the meetings and the global nature of colleges means that more often than not, the class will have students missing. To remedy this, Zoom, the platform which most universities use for synchronizing their meetings, offers a function where it records the meetings, and automatically generates a transcript for the meeting. While this is a good stopgap to ensure that students can access lectures in a timely manner, it doesn't solve the core problem that students might not have the time to fully review a lecture in-depth. As such, we seek to answer the question: **is it possible to create an effective text summarizer that is specialized towards the long-form formats of lectures and presentations?**

At a first glance this question seems rather simple to answer, being that text summarization is a problem that has been studied before, and there are papers that have created effective summarizers using models. However, from our own research, not much has been done for trying to summarize longer-form text, because the task becomes harder the more text there is to try and summarize. While this tends towards impossibility with large, jargon-heavy works such as research papers and technical novel chapters, the

length and complexity of university presentations and lectures is actually a perfect sweet-spot in terms of summarizing, which we will explore further in this paper.

Related Work

Text Summarization Techniques: A Brief Survey¹

This paper is the one we used the most as a reference point for our own work. It outlines a large series of text summarization techniques, and goes over each one in enough detail to get a good idea of the benefits and limitations of any given approach to text summarization.

Specifically, from this paper we got the approach of using Extractive Summarization as the main form of summarization used in our own summarizer model.

In addition to the type of summarization, we also obtained a good idea for evaluation metrics for our model. Specifically, the ROGUE-n and ROGUE-L evaluation metrics were chosen as ideal due to the fact that they're based on recall rather than precision and f1. This was necessary because the precision and f1 are naturally going to be low for a summary evaluation, since the summary does not contain much of the original text, and precision/f1 are context independent, so they do not actually take into account the fact that our task wants to minimize the amount of words we pick up from the text.

Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation²

This paper was the second most important one for generating our models. Not all of it was relevant to the summarization task, in particular the entire section about image captioning was not relevant to the project, however the task of natural language generation was highly important, since that's the basis of actually generating a summary.

The subsection 'NLG Tasks' was the most referenced part of this paper, the idea of content determination, text structuring, and sentence aggregation were the most important parts of creating a summary, and using lexicalisation to find out which jargon in a text block was important to include in a summary and which jargon isn't was helpful when dealing with more technical text blocks.

Finally, for this paper, the 'Evaluation' section was referenced. This served to strengthen the reliance on ROGUE-n/ROGUE-L metrics that were obtained from the paper "Text summarization techniques", as n-gram overlap was unreliable, and the fact that ROGUE-L measured longest common subsequences was relevant to finding out similarities between the summary and the base text block.

¹ Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (n.d.). Text Summarization Techniques: A Brief Survey. Retrieved October, 2020, from <https://arxiv.org/pdf/1707.02268.pdf>

² Gatt, A., & Krahmer, E. (2018, January 29). Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation. Retrieved November 03, 2020, from <https://arxiv.org/abs/1703.09902>

SINGLE DOCUMENT AUTOMATIC TEXT SUMMARIZATION USING TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF)³

Given that Extractive Summarization was one of the techniques that we were most seriously considering for creating the automatic summarizer, TF-IDF became a natural choice for the algorithm used to generate the summaries. This paper in particular is more of a deep dive into the use of the TF-IDF algorithm pertaining to single document summarization. This lines up nicely with the mission statement of our research, which involves single document summarization of a long-form text block. As such, this paper became rather important to read and review since the techniques and explanations they used for TF-IDF were highly pertinent to the summarization task that we were going for.

Finally, the flowcharts outlined in the paper were very helpful in getting an idea of how to implement TF-IDF since they were very good descriptors of the logical processes involved. For example, the following flowchart was used multiple times as a reference for creating the summarization program's logical skeleton.

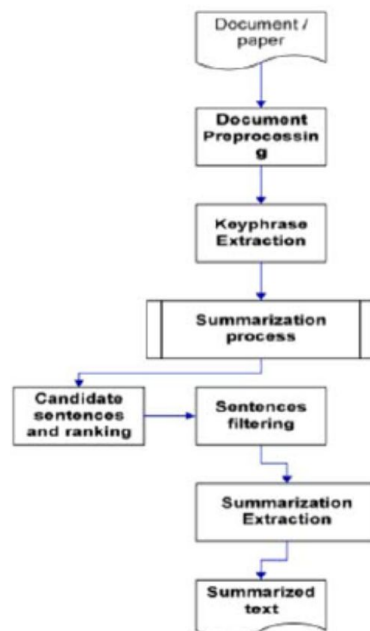


Figure 1: Summarization extraction program flowchart from “Single document... TF-IDF”

³ Christian, H., Agus, M. P., & Suhartono, D. (2016, December). SINGLE DOCUMENT AUTOMATIC TEXT SUMMARIZATION USING TERM ... Retrieved November 12, 2020, from http://research-dashboard.binus.ac.id/uploads/paper/document/publication/Proceeding/ComTech/Vol.%2007%20No.%204%20December%202016/005_TI_Hans_rev1.pdf

Comparison of Multi Document Summarization Techniques Single V/S Multi Document Summarization⁴

Finally, this paper was a good look into additional techniques used for summarizing text. The MEAD extraction framework is a publicly available one, and was considered when thinking about the actual tools that were going to be used for summarization. While it wasn't used in the final version of the text summarizer that we implemented, the evaluation techniques used in this paper were important because they gave us a good idea of how to go about accurately evaluating our model with the metrics that we had chosen.

Data

Before going into the methods used and the results, the dataset used must be discussed, because the actual data that the model is trained on plays a large part in the summarization outcomes.

Initially, the dataset was going to consist of Zoom lectures' auto-generated transcripts. Every Zoom lecture that has this option enabled allows people with access to download the transcripts as a text file.

In the final version of the dataset, Zoom lectures were not included. This is because the University of Massachusetts Amherst Zoom lectures do not have a downloadable transcript option, and copying directly from the chat log creates a text log with far too much formatting to clear programmatically, and manually cleaning up massive amounts of text for many lectures would not only be immensely time consuming, but also impractical with only two people. As such, we instead opted for the transcripts of TED talks as our dataset. These transcripts have a similar format to Zoom lectures, and contain similar amounts of technical jargon. While the transcripts themselves are shorter than a standard Zoom lecture transcript (in terms of word count), the techniques used to summarize a TED talk can be upscaled to a Zoom lecture rather easily. The TED dataset used was manually gathered by going to various TED talk pages and downloading the txt file corresponding to the actual talk. In this way, 50 different talks were downloaded and combined form up the core dataset for our summarizer. While 50 talks may not seem like much, it corresponds to a very large amount of text, since each transcript contains at least 3000 words, and each talk has a large amount of diverse technical jargon. As such, it provides a robust and diverse set of text for the summarizer to train on. This is reflected in the actual summaries produced since they tend to contain more of the jargon in its contextual usage.

Finally, a few Zoom lectures were actually downloaded and hand-cleaned. The reasoning for this is because a test set was needed to ensure that the summarizer works on even larger text blocks, such as those found in Zoom lectures. The size difference between an average TED talk and an average 1 hour 15 minutes Zoom lecture is about 1:5 in terms of word count. Likely because a university lecturer is more likely to take pauses while explaining, so the word count difference is not that much when compared to the actual length difference of the two talks, with TED talks averaging at around 10 minutes, a 1:7 ratio as compared to the word count's 1:5.

⁴ Nedunchelian, R., Muthukumaraswamy, R., & Saranathan, E. (2011). Comparison of multi document summarization techniques. *International Journal of Computer Applications*. 11(3), 155-160.

Method

For our method, we tried to use different models. Some of the models that we used were the pytextrank algorithm, the BERT model and TF-IDF algorithm. We first tried to use the pytextrank algorithm but that did not yield very good results with the type of data we had. Next we tried to use the BERT model but that didn't do much either. So hence we decided to go for the TF-IDF approach and that's what we used to get our results.

TF-IDF is short for Term Frequency * inverse Document Frequency. This algorithm is made up of 2 separate algorithms. The first is the term frequency, which is defined as how often a word appears in a document divided by the total number of words in the document. The second algorithm is the inverse document frequency which is how unique or rare a word is. The formula of inverse document frequency is:

$$IDF(t) = \ln(\text{Total number of words} / \text{number of } t \text{ words})$$

Thus the TF-IDF is just the product of these 2 algorithms. We use the same concept where we generate our sentence tokens, remove all our stopwords, create the frequency matrix for each word in the sentence. Then we generate the term frequency matrix and the IDF matrix and we combine both to generate the TF-IDF values. Then we score the sentence based on these values. After that, we calculate the average score and generate our summary based on the threshold we apply. The threshold is used to multiply it by the average score and then we choose the sentences which have a score better than the product of the threshold and the average score. Hence, our only hyperparameter is the threshold value which we set and is user dependent. How the threshold works is If the user wants a concise summary, you set a higher threshold but if the user wants their summary to be a bit longer, we set a lower threshold. Therefore, higher the threshold value, the more concise the summary.

To evaluate our model, we went with the ROGUE metric. The ROGUE(Recall Oriented Understudy for Gisting Evaluation) is a set of metrics used for evaluation of automatic text summarization. It consists of 5 evaluation metric but we are only concerned with 3, which are:

- ROGUE-1: Measures overlap of unigrams between summary and reference
 - ROGUE-2: Measures overlap of bi-grams between summary and reference
 - ROGUE-L: Computes the longest subsequence between the reference and the summary.
- Each sentence is considered as a sequence of words.

Each metric measures its own precision, recall and f1 score and reports them. So for our evaluation we took 50 Ted talks and measured the evaluation of our model with each of these 3 metrics.

And before we did any of this, it was important we preprocess our ted talks transcripts because there was a lot of punctuation and a lot of action words like (*applause*) or (*laughter*). So we had to get rid of all this additional noise in our data which could hinder the summarization. Once this was done, we processed to use the TF-IDF algorithm to generate our summarization.

Results

As was outlined earlier, the threshold for ROGUE-n and ROGUE-L was set to values of 1, 1,1 and 1.2, using these, a summary was performed on each transcript in the data set, and evaluation metrics were gathered for each one.

Since there are 50 data points, and 9 possible permutations of threshold/metric (with the 3 metrics being ROGUE-1, ROGUE-2, and ROGUE-L), this would make 9 tables of size 50 each. However, for the sake of brevity most of this data is being excluded from the report. Instead, the average accuracy scores for an average word count will be reported for each of these metrics.

Threshold	Avg Word Count	Avg Precision	Avg Recall	Avg F1 Score
1	3471.412	0.162315	0.999278	0.277682
1.1	3471.412	0.124673	0.999209	0.220693
1.2	3471.412	0.099802	0.998965	0.180822

Table 1: ROGUE-1 metrics for all 3 thresholds

Threshold	Avg Word Count	Avg Precision	Avg Recall	Avg f1 score
1	3471.412	0.151616	0.934514	0.259419
1.1	3471.412	0.114783	0.922222	0.20323
1.2	3471.412	0.090816	0.912072	0.164582

Table 2: ROGUE-2 metrics for all 3 thresholds

Thresholds	Avg Word Count	Avg Precision	Avg Recall	Avg F1 Score
1.1	3471.412	0.27158	0.998461	0.424343
1.2	3471.412	0.221731	0.998212	0.360905
1.3	3471.412	0.185984	0.9983	0.312249

As can be seen from the ROGUE-1 metrics, the precision scores and f1 scores for the summaries are low on average. However this is an expected result due to how precision and f1 are measured. What's really relevant here is the very high recall scores overall. This is significant because recall is going to be our

primary way of measuring the efficacy of our model. This is because summarization is a primarily recall based task, since we're looking for the amount of relevant words that show up in our document, as opposed to how similar it is to the document overall. Our ROGUE-2 metric backs this up, showing that the 2-gram combinations also have a high recall, which means that the 2-grams showing up in the summary tend to be used a lot within the summary itself, indicating relevancy.

Finally, the ROGUE-L metric is arguably the most important metric of the 3. In particular because it measures the longest common subsequence between the summary and the actual text. We also notice that the ROGUE-L metric tends to have a higher precision and f1 score than the ROGUE-1 and ROGUE-2 ones. This likely indicates that the longest subsequences of characters in the summary are more accurate to the ones found in the test. I.e. important quotes are being lifted from the text and dropped into the summary where relevant (as indicated by the recall scores still being rather high).

Here is one of the summaries we had generated based on the text.

Text:

good morning. how are you? it's been great, hasn't it? i've been blown away by the whole thing. in fact, i'm leaving. there have been three themes running through the conference which are relevant to what i want to talk about. one is the extraordinary evidence of human creativity in all of the presentations that we've had and in all of the people here. just the variety of it and the range of it. the second is that it's put us in a place where we have no idea what's going to happen, in terms of the future. no idea how this may play out. i have an interest in education. actually, what i find is everybody has an interest in education. don't you? i find this very interesting. if you're at a dinner party, and you say you work in education — actually, you're not often at dinner parties, frankly. if you work in education, you're not asked, and you're never asked back, curiously. that's strange to me. but if you are, and you say to somebody, you know, they say, "what do you do?" and you say you work in education, you can see the blood run from their face. they're like, "oh my god." you know, "why me?" "my one night out all week." but if you ask about their education, they pin you to the wall. because it's one of those things that goes deep with people, am i right? like religion, and money and other things. so i have a big interest in education, and i think we all do. we have a huge vested interest in it, partly because it's education that's meant to take us into this future that we can't grasp. if you think of it, children starting school this year will be retiring in 2065. nobody has a clue, despite all the expertise that's been on parade for the past four days, what the world will look like in five years' time. and yet we're meant to be educating them for it. so the unpredictability, i think, is extraordinary. and the third part of this is that we've all agreed, nonetheless, on the really extraordinary capacities that children have — their capacities for innovation. i mean, sirena last night was a marvel, wasn't she? just seeing what she could do. and she's exceptional, but i think she's not, so to speak, exceptional in the whole of childhood. what you have there is a person of extraordinary dedication who found a talent. and my contention is, all kids have tremendous talents. and we squander them, pretty ruthlessly. so i want to talk about education and i want to talk about creativity. my contention is that creativity now is as important in education as literacy, and we should treat it with the same status. thank you. that was it, by the way. thank you very much. so, 15 minutes left. well, i was born... no. i heard a great story recently — i love telling it — of a little girl who was in a drawing lesson. she was six, and she was at the back, drawing, and the teacher said this girl hardly ever paid attention, and in this drawing lesson, she did. the teacher was fascinated. she went over to her, and she said, "what are you drawing?" and the girl said, "i'm drawing a picture of god." and the teacher said, "but nobody knows what god looks like." and the girl said, "they will, in a minute." when my son was four in england — actually, he was four everywhere, to be honest. if we're being strict about it, wherever he went, he was four that year. he was in the nativity play. do you remember the story? no, it was big, it was a big story. mel gibson did the sequel, you may have seen it. "nativity ii." but james got the part of joseph, which we were thrilled about. we considered this to be one of the lead parts. we had the place crammed full of agents in t-shirts: "james robinson is joseph!" he didn't have to speak, but you know the bit where the three kings come in? they come in bearing gifts, gold, frankincense and myrrh. this really happened. we were sitting there and i think they just went out of sequence, because we talked to the little boy afterward and we said, "you ok with that?" and he said, "yeah, why? was that wrong?" they just switched. the three boys came in, four-year-olds with tea towels on their heads, and they put these boxes down, and the first boy said, "i bring you gold." and the second boy said, "i bring you myrrh." and the third boy said, "frank sent this. 'what these things have in common is that kids will take a chance. if they don't know, they'll have a go. am i right?' they're not frightened of being wrong. i don't mean to say that being wrong is the same thing as being creative. what we do know is, if you're not prepared to be wrong, you'll never come up with anything original — if you're not prepared to be wrong. and by the time they get to be adults, most kids have lost that capacity. they have become frightened of being wrong, and we run our companies like this. we stigmatize mistakes. and we're now running national education systems where mistakes are the worst thing you can make. and the result is that we are educating people out of their creative capacities. picasso once said this, he said that all children are born artists. the problem is to remain an artist as we grow up. i believe this passionately, that we don't grow into creativity, we grow out of it. or rather, we get educated out of it. so why is this? i lived in stratford-on-avon until about five years ago. in fact, we moved from stratford to los angeles. so you can imagine what a seamless transition that was. actually, we lived in a place called snitterfield, just outside stratford, which is where shakespeare's father was born. are you struck by a new thought? i was. you don't think of shakespeare having a father, do you? do you? because you don't think of shakespeare being a child, do you? shakespeare being seven? i never thought of it. i mean, he was seven at some point. he was in somebody's english class, wasn't he? how annoying would that be? "must try harder." being sent to bed by his dad, you know, to shakespeare, "go to bed, now! and put the pencil down." "and stop speaking like that." "it's confusing everybody." anyway, we moved from stratford to los angeles, and i just want to say a word about the transition. my son didn't want to come. i've got two kids; he's 21 now, my daughter's 16. he didn't want to come to los angeles. he loved it, but he had a girlfriend in england. this was the love of his life, sarah. he'd known her for a month. mind you, they'd had their fourth anniversary, because it's a long time when you're 16. he was really upset on the plane, he said, "i'll never find another girl like sarah." and we were rather pleased about that, frankly — because she was the main reason we were leaving the country. but something strikes you when you move to america and travel around the world: every education system on earth has the same hierarchy of subjects. every one. doesn't matter where you go. you'd think it would be otherwise, but it isn't. at the top are mathematics and languages, then the humanities, and at the bottom are the arts. everywhere on earth. and in pretty much every system too, there's a hierarchy within the arts. art and music are normally given a higher status in schools than drama and dance. there isn't an education system on the planet that teaches dance everyday to children the way we teach them mathematics. why? why not? i think this is rather important. i think math is very important, but so is dance. children dance all the time if they're allowed to, we all do. we all have bodies, don't we? did i miss a meeting? truthfully, what happens is, as children grow up, we start to educate them progressively from the waist up. and then we focus on their heads. and slightly to one side. if you were to visit education, as an alien, and say "what's it for, public education?" i think you'd have to conclude, if you look at the output, who really succeeds by this, who does everything that they should, who gets all the brownie points, who are the winners — i think you'd have to conclude the whole purpose of public education throughout the world is to produce university professors. isn't it? they're the people who come out the top. and i used to be one, so there. and i like university professors, but you know, we shouldn't hold them up as the high-water mark of all human achievement. they're just a form of life, another form of life. but they're rather curious, and i say this out of affection for them. there's something curious about professors in my experience — not all of them, but typically, they live in their heads. they live up there, and slightly to one side. they're disembodied, you know, in a kind of literal way. they look upon their body as a form of transport for their heads. don't

they? it's a way of getting their head to meetings. if you want real evidence of out-of-body experiences, get yourself along to a residential conference of senior academics, and pop into the discotheque on the final night. and there, you will see it. grown men and women writhing uncontrollably, off the beat. waiting until it ends so they can go home and write a paper about it. our education system is predicated on the idea of academic ability. and there's a reason. around the world, there were no public systems of education, really, before the 19th century. they all came into being to meet the needs of industrialism. so the hierarchy is rooted on two ideas. number one, that the most useful subjects for work are at the top. so you were probably steered benignly away from things at school when you were a kid, things you liked, on the grounds that you would never get a job doing that. is that right? don't do music, you're not going to be a musician; don't do art, you won't be an artist. benign advice — now, profoundly mistaken. the whole world is engulfed in a revolution. and the second is academic ability, which has really come to dominate our view of intelligence, because the universities designed the system in their image. if you think of it, the whole system of public education around the world is a protracted process of university entrance. and the consequence is that many highly-talented, brilliant, creative people think they're not, because the thing they were good at at school wasn't valued, or was actually stigmatized. and i think we can't afford to go on that way. in the next 30 years, according to unesco, more people worldwide will be graduating through education than since the beginning of history. more people, and it's the combination of all the things we've talked about — technology and its transformation effect on work, and demography and the huge explosion in population. suddenly, degrees aren't worth anything. isn't that true? when i was a student, if you had a degree, you had a job. if you didn't have a job, it's because you didn't want one. and i didn't want one, frankly. but now kids with degrees are often heading home to carry on playing video games, because you need an ma where the previous job required a ba, and now you need a phd for the other. it's a process of academic inflation. and it indicates the whole structure of education is shifting beneath our feet. we need to radically rethink our view of intelligence. we know three things about intelligence. one, it's diverse. we think about the world in all the ways that we experience it. we think visually, we think in sound, we think kinesthetically. we think in abstract terms, we think in movement. secondly, intelligence is dynamic. if you look at the interactions of a human brain, as we heard yesterday from a number of presentations, intelligence is wonderfully interactive. the brain isn't divided into compartments. in fact, creativity — which i define as the process of having original ideas that have value — more often than not comes about through the interaction of different disciplinary ways of seeing things. by the way, there's a shaft of nerves that joins the two halves of the brain called the corpus callosum. it's thicker in women. following off from helen yesterday, this is probably why women are better at multi-tasking. because you are, aren't you? there's a raft of research, but i know it from my personal life. if my wife is cooking a meal at home — which is not often, thankfully. no, she's good at some things, but if she's cooking, she's dealing with people on the phone, she's talking to the kids, she's painting the ceiling, she's doing open-heart surgery over here. if i'm cooking, the door is shut, the kids are out, the phone's on the hook, if she comes in i get annoyed. i say, "terry, please, i'm trying to fry an egg in here." "give me a break." actually, do you know that old philosophical thing, if a tree falls in a forest and nobody hears it, did it happen? remember that old chestnut? i saw a great t-shirt recently, which said, "if a man speaks his mind in a forest, and no woman hears him, is he still wrong?" and the third thing about intelligence is, it's distinct. i'm doing a new book at the moment called "epiphany," which is based on a series of interviews with people about how they discovered their talent. i'm fascinated by how people got to be there. it's really prompted by a conversation i had with a wonderful woman who maybe most people have never heard of, gillian lynne. have you heard of her? some have. she's a choreographer, and everybody knows her work. she did "cats" and "phantom of the opera." she's wonderful. i used to be on the board of the royal ballet, as you can see. anyway, gillian and i had lunch one day and i said, "how did you get to be a dancer?" it was interesting. when she was at school, she was really hopeless. and the school, in the '30s, wrote to her parents and said, "we think gillian has a learning disorder." she couldn't concentrate; she was fidgeting. i think now they'd say she had adhd. wouldn't you? but this was the 1930s, and adhd hadn't been invented at this point. it wasn't an available condition. people weren't aware they could have that. anyway, she went to see this specialist. so, this oak-paneled room, and she was there with her mother, and she was led and sat on this chair at the end, and she sat on her hands for 20 minutes while this man talked to her mother about the problems gillian was having at school. because she was disturbing people; her homework was always late; and so on, little kid of eight. in the end, the doctor went and sat next to gillian, and said, "i've listened to all these things your mother's told me, i need to speak to her privately. wait here. we'll be back; we won't be very long," and they went and left her. but as they went out of the room, he turned on the radio that was sitting on his desk. and when they got out, he said to her mother, "just stand and watch her." and the minute they left the room, she was on her feet, moving to the music. and they watched for a few minutes and he turned to her mother and said, "mrs. lynne, gillian isn't sick; she's a dancer. take her to a dance school." i said, "what happened?" she said, "she did. i can't tell you how wonderful it was. we walked in this room and it was full of people like me. people who couldn't sit still. people who had to move to think," who had to move to think. they did ballet, they did tap, jazz; they did modern; they did contemporary. she was eventually auditioned for the royal ballet school; she became a soloist; she had a wonderful career at the royal ballet. she eventually graduated from the royal ballet school, founded the gillian lynne dance company, met andrew lloyd webber. she's been responsible for some of the most successful musical theater productions in history, she's given pleasure to millions, and she's a multi-millionaire. somebody else might have put her on medication and told her to calm down. what i think it comes to is this: al gore spoke the other night about ecology and the revolution that was triggered by rachel carson. i believe our only hope for the future is to adopt a new conception of human ecology, one in which we start to reconstitute our conception of the richness of human capacity. our education system has mined our minds in the way that we strip-mine the earth: for a particular commodity. and for the future, it won't serve us. we have to rethink the fundamental principles on which we're educating our children. there was a wonderful quote by jonas salk, who said, "if all the insects were to disappear from the earth, within 50 years all life on earth would end. if all human beings disappeared from the earth, within 50 years all forms of life would flourish." and he's right. what ted celebrates is the gift of the human imagination. we have to be careful now that we use this gift wisely and that we avert some of the scenarios that we've talked about. and the only way we'll do it is by seeing our creative capacities for the richness they are and seeing our children for the hope that they are. and our task is to educate their whole being, so they can face this future. by the way — we may not see this future, but they will. and our job is to help them make something of it. thank you very much.

Generated Summary:

good morning. i've been blown away by the whole thing. just the variety of it and the range of it. don't you? i find this very interesting. that's strange to me. ""my one night out all week. and yet we're meant to be educating them for it. just seeing what she could do. what you have there is a person of extraordinary dedication who found a talent. and my contention is, all kids have tremendous talents. the teacher was fascinated. he was in the nativity play. do you remember the story? no, it was big, it was a big story. mel gibson did the sequel, you may have seen it. "nativity ii." we considered this to be one of the lead parts. this really happened. was that wrong?" they just switched. am i right? they're not frightened of being wrong. they have become frightened of being wrong. and we run our companies like this. we stigmatize mistakes. the problem is to remain an artist as we grow up. are you struck by a new thought? i was. do you? shakespeare being seven? i never thought of it. "must try harder. and put the pencil down. ""and stop speaking like that. ""it's confusing everybody. my son didn't want to come. he loved it, but he had a girlfriend in england. every one. doesn't matter where you go. everywhere on earth. why? why not? i think this is rather important. we all have bodies. don't we? and then we focus on their heads. isn't it? they're the people who come out the top. they're just a form of life, another form of life. they look upon their body as a form of transport for their heads. don't they? and there's a reason. they all came into being to meet the needs of industrialism. is that right? benign advice — now, profoundly mistaken. isn't that true? it's a process of academic inflation. and it indicates the whole structure of education is shifting beneath our feet. we think about the world in all the ways that we experience it. we think visually, we think in sound, we think kinesthetically. we think in abstract terms, we think in movement. secondly, intelligence is dynamic. the brain isn't divided into compartments. it's thicker in women. ""give me a break. remember that old chestnut? i'm fascinated by how people got to be there. have you heard of her? she did "cats" and "phantom of the opera." she's wonderful. it was interesting. i think now they'd say she had adhd. wouldn't you? wait here. take her to a dance school. who had to move to think. they did ballet, they did tap, jazz; they did modern; they did contemporary. and our job is to help them make something of it. thank you very much.

Analysis

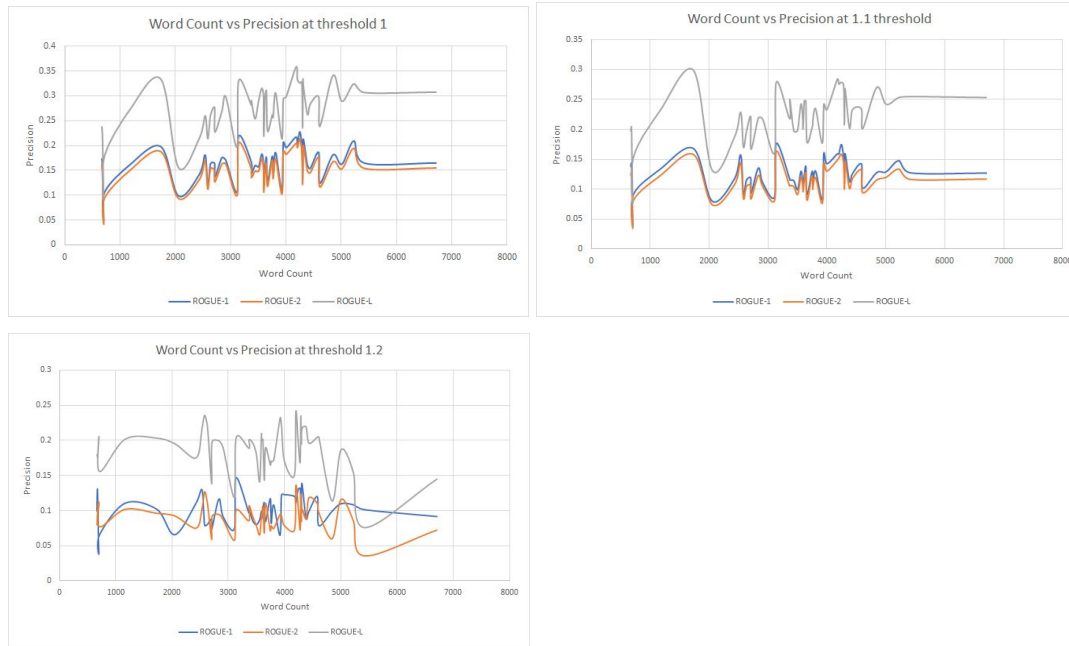


Figure 2, 3, 4: Word Count vs Precision at thresholds 1, 1.1, 1.2

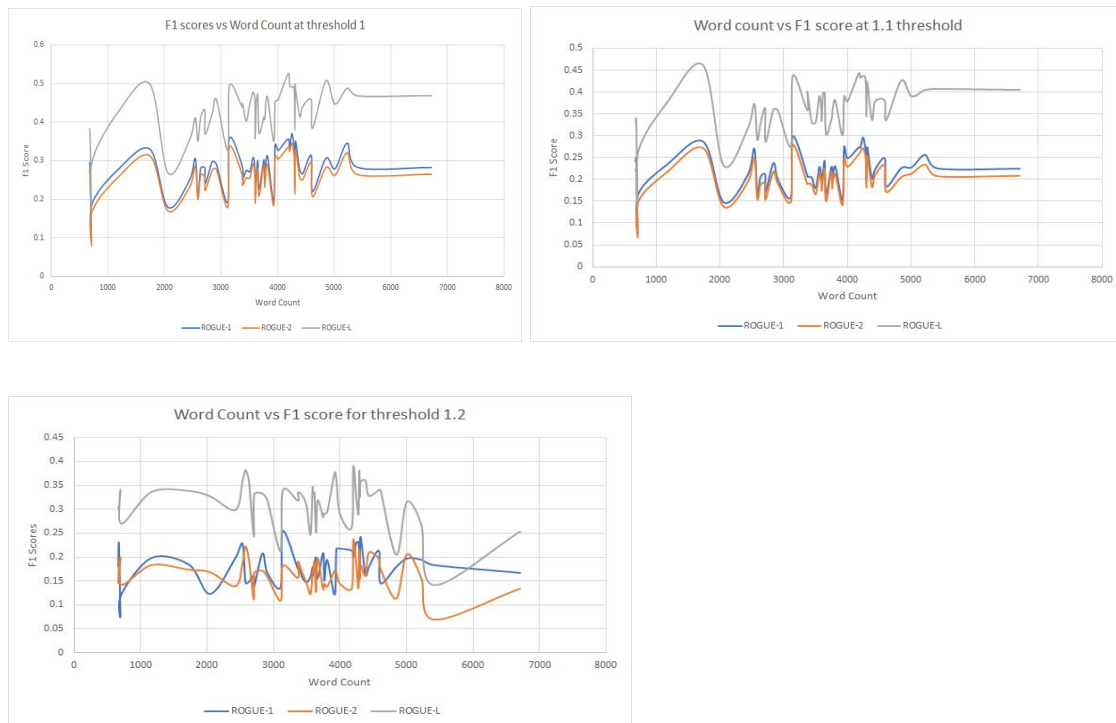


Figure 5, 6, 7: Word Count vs f1 score at thresholds 1, 1.1, 1.2

If we take the average the F1 scores for all the threshold values for each metric, we see that our the average scores for F1 for each metric are:

- ROGUE-1 \approx 0.3
- ROGUE-2 \approx 0.33 - 0.35
- ROGUE-L \approx 0.5

Now the reason ROGUE-1 and ROGUE-2 are pretty low is because now all words from the reference get generated into the summary. So it makes sense that they are low, considering that we are doing extractive summarization. But the ROGUE-L metric is the most important one and considering that we only generated 3-4 line summaries, it is pretty decent that we got the average F1 score of about 0.5. If we went for an even smaller threshold, We could have generated an even larger summary(around 8-9 lines) and our ROGUE-L F1 score, as well as the F1 scores of ROGUE-1 and ROGUE-2, would have definitely improved. We can see this because in the graphs above for word count vs F1-score, we see that with higher thresholds, the scores were lower.

Another way to have done the analysis and a very simple way is to ask Humans to assess its quality. Humans always have the best brains and when it comes to grammar and precision and accuracy of the summary, only the humans do it best. So we also asked our friends to assess how well they thought the summaries were and they said they were able to get quite an understanding of our summaries and could figure out what it was trying to convey. They indicated that the grammar was a bit off and some parts were unclear but for the most part it was understandable.

At the end, this is still a summary no matter how large or small the summary is. A summary is never going to tell you every single fact that is present in the reference. A summary, as defined, is a brief statement or account of the major points of a document. So to be able to even achieve a F1 score of 0.5 means that we got 50 percent of the entire summary, including all the major points(which were set based on the sentence score values) was a promising result.

Discussion and Future Work

Overall, the summarizer we produced is a rather reliable model that is able to handle somewhat long-form text blocks with a good recall. This is confirmed by the ROGUE scores indicating not only a high recall, but a relatively high precision and F1 with the context that summaries tend to differ drastically from the original texts.

As for the summaries themselves, for the most part they were coherent and (from manual checking) rather apt summaries of the transcripts they come from. In some cases they were even more descriptive than the summaries that TED provides for their talks. However, as with any text generation, there were some problems with the grammar when generating text. To an extent this is unavoidable, because too much training can cause the data to overfit and produce summaries that are just sentences directly ripped from the original text, however any amount of training that isn't overfitting will produce irregularities as the program tries to interpolate text, which might not necessarily fit with the laws of grammar.

That being said, the project was mostly a success, and we did answer the question as to whether it is possible to create an effective summarizer for long-form text, the answer being a resounding “Yes”. As for where this can go in the future, the next step is to try and upscale this even further, attempting to use it on even longer blocks of text to see where it can go until. This could be useful for either longer speeches or interviews, like interviews with politicians and other figures of significance, or longer form text such as the transcripts of video essays and podcasts. The applications of a long-form text summarizer are really varied due to how much media exists in which people just talk for a long period of time, and creating a framework that is applicable to the shorter end of the long-form spectrum opens up possibilities of editing or using it for even longer texts.

Bibliography

1. Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (n.d.). Text Summarization Techniques: A Brief Survey. Retrieved October, 2020, from <https://arxiv.org/pdf/1707.02268.pdf>
2. Gatt, A., & Krahmer, E. (2018, January 29). Survey of the State of the Art in Natural Language Generation: Core tasks, applications and evaluation. Retrieved November 03, 2020, from <https://arxiv.org/abs/1703.09902>
3. Christian, H., Agus, M. P., & Suhartono, D. (2016, December). SINGLE DOCUMENT AUTOMATIC TEXT SUMMARIZATION USING TERM ... Retrieved November 12, 2020, from http://research-dashboard.binus.ac.id/uploads/paper/document/publication/Proceeding/ComTech/Vol.%2007%20No.%204%20December%202016/005_TI_Hans_rev1.pdf
4. Nedunchelian, R., Muthukumaraswamy, R., & Saranathan, E. (2011). Comparison of multi document summarization techniques. International Journal of Computer Applications. 11(3), 155-160.